

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326224279>

Text-to-Speech Device for Visually Impaired People

Article in International Journal of Pure and Applied Mathematics · July 2018

CITATION

1

READS

8,516

1 author:



[Shirly Edward](#)

SRM Institute of Science and Technology

26 PUBLICATIONS 50 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Implementation of smart sensor interface network for water quality monitoring in industry using IoT [View project](#)

Text-To-Speech Device for Visually Impaired People

Shirly Edward.A¹

Assistant Professor

Dept of ECE,

SRMIST, Vadapalani,
ChennaiJothimani.A²

Assistant Professor

Dept of ECE,

SRMIST, Vadapalani,
ChennaiJayaprakash.V³

U.G Student

Dept of ECE,

SRMIST, Vadapalani,
ChennaiJoe Benhur Xavier.F⁴

U.G Student

Dept of ECE,

SRMIST, Vadapalani,
Chennai

Abstract - People who suffer from low vision, sight and visual impairment are not able to see words and letters in ordinary newsprint, books and magazines clearly. This can make the reading process difficult which can disturb the learning process and slow the person's intelligence development. Therefore, a device is needed to help them read. So we had developed one such device that can scan and read any kind of text by changing it to voice message. The purpose of this device is to process the input Image, pdf, Documents, Textbooks, and News Papers as input into a voice as output. Each Module for image processing and voice processing are present in the device. It also has the ability to play and stop the output while reading. It has less error rate and less processing time and cost efficiency. Raspberry pi 3 was used to develop the device. This device actually acts as an artificial eye to visually impaired people. It doesn't need any human supervision.

vision. So, that they can easily use this device without having to ask for help from others and they can utilize this device for academic and intelligence ability.

I.Introduction

Based on the survey by World Health organization in 2010, total population in India is 1181.4 million out of which people who suffer with blindness, low vision and visual impairment are 152.238 Million[1]. Figure 1 shows the number of people who are blind, with low vision and visually impaired (in thousands) per million population.

According to Dr.Bjorn, impaired vision can have negative effects on learning and social interaction. It can affect the natural development of intelligence and academic ability, social, and profession [2]. People who are visually impaired cannot be recovered with the help of glasses. This causes people with low vision, they cannot even see the normal printed paper. They can only see if the sizes of the characters or letters are big enough. This condition impacted the length of the reading process and made the eyes tired. To help improve the quality of life for people with low vision a tool to read the article is needed. The rate of vision impairment can vary in each individual with low vision. Therefore a device developed in this work utilized other sensory function in receiving information from a text. The device is specifically designed for the people with low

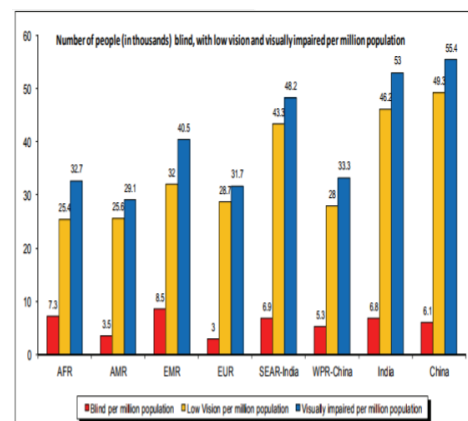


Figure 1. Ratio of people with visual impairment
(courtesy: WHO, 2010)

II. Hardware and Software

A. Raspberry Pi 3

The Raspberry Pi is a small computer which can be fit in the palm of one's hand. It runs on Linux and also with few other low-power operating systems. It was created by the Raspberry Pi Foundation. Raspbian is the official operating system of Raspberry Pi. Other operating systems exist, but they're mostly made for specific projects. It has Broadcom BCM2837 64bit Quad Core Processor powered Single Board Computer running at 1.2GHz with 1GB RAM. The Raspberry Pi has gone through several iterations since its launch in 2012. The latest version is the Raspberry Pi Model 3. It is available for 3200 INR. It has a ceaseless computer that includes an HDMI output, up to four USB ports, Wi-Fi, and Bluetooth.

B. Raspberry Pi Camera Module V2.1

The Raspberry Pi has a Camera Module V2 with high quality and 8 megapixel Sony IMX219 image sensor which is custom designed and acts as add-on board for Raspberry Pi, with fixed focus lens. It is capable of acquiring 3280 x 2464 pixel static images, and also supports 1080p30, 720p60 and 640x480p60/90 video. The add-on board attaches to Pi via one of the small sockets present on the board upper surface and uses the dedicated CSI interface, designed especially for interfacing to cameras. The size of the board is 25mm x 23mm x 9mm. It also weighs just over 3g, making it perfect for mobile or other applications where size and weight are not of much importance. Figure 2 shows the image of Raspberry Pi 3 class B module.



Figure 2. Raspberry pi 3 Class B

C. Python IDLE

Python, created by Guido van Rossum is a high-level programming language mainly used for general-purpose programming. This language gives constructs intended to enable writing clear programs on both a little and vast scale. Python highlights a dynamic sort framework and automatic memory management and supports numerous programming paradigms, including object-oriented, functional and imperative programming. It has a huge and extensive standard library. Python is a wonderful and very useful programming language. It is easy to use and with Raspberry Pi it lets us to convert one's project into real-time. Python syntax is very clean, with an emphasis on readability and uses Standard English keywords. Start by opening IDLE. The easiest introduction to Python is through IDLE, a Python Integrated DeveLopment Environment.

D. System Specification and Design

The device is designed based on the following restrictions:

- Range of reading distance is 15-30 cm.
- Character size is minimum of 8 pt.
- Maximum size of reading material can be varied.
- Maximum tilt of the text line is 5 degrees from the Vertical
- Type of characters includes Roman, Egyptian or Sans Serif types.

The module is designed in such a way that there is no physical equipment or stand like structure is used, to carry the pi cam module, as it is placed using two L-clamps over the encasing of the board. The pi camera lens is adjusted, in order to acquire the script sharply. The distance between the cam module and script is between 15 to 30cm, the minimal distance that a human eye needs to read a script.

3. System Architecture

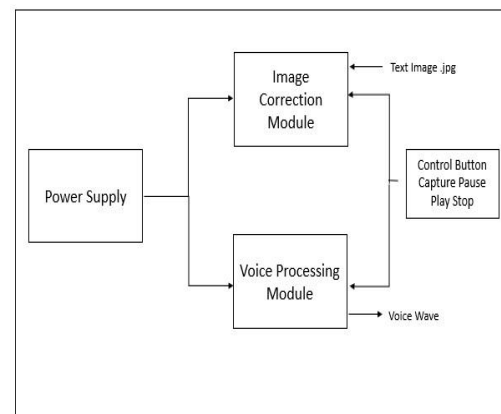


Figure 3 Architecture of the Device

Figure 3 shows the architecture of the device. It mainly consists of Image Correction Module and TTS correction and Voice Processing Module. Each of these modules are elaborated in the consequent sections.

A. Image Correction module

(a) Gray Scaling

It is a process of converting a digital or pixel image into a gray scaled image [3]. Each value of pixel is defined as single sample as it carries only the information about intensity. These are composed exclusively of gray shades, varying from weakest intensity i.e., black to white the strongest intensity ranging from 0 to 255.

(b) Binarization

It is a process of converting a grey scaled image into a binary image. It basically consists of black and white. Binarization comes with thresholding .Every pixel with a value greater than 170 turns white (gets the value of 255) and every pixel with a value lesser than 170 becomes black (gets the value of 0).

(c) **Image Processing module using Optical Character Recognition**

This module consists of OCR or Optical Character Recognition. It targets typewritten text, one glyph or character at a time. OCR utilizes optical mechanism to automatically recognize the characters, this technology imitate the ability of the human senses of sight, where the camera replaces the eye and processing of image is done in the computer as a substitute for the human brain. OCR engine required state and initial steps in order to get the best input of OCR to reduce the disability of this OCR engine. Setup state is well adapted to the specifications of the desired initial device. So that the desired output of this processing has a minimum error rate is also a short processing time. This module does not change the OCR algorithm, but gives additional state to get the best input of OCR. It is generally an "offline" process, which analyses a static document.

Tesseract OCR is a type of OCR engine with matrix matching. The selection of Tesseract engine is due to the wide acceptance in the world because of flexibility and extensibility of these engines and the fact that many researchers are actively developing this OCR engine [4]. Defects in machines such as distortion at the edges and the dim light effect make it difficult for most OCR engines to get high accuracy text

Tesseract OCR Implementation

The input image captured by the camera has a size of 8 MP or 215 ppi (pixels per inch) [5]. As per the specifications of the Tesseract OCR engine, 20 pixels uppercase letters is the minimum character size that can be read. Tesseract OCR accuracy will decrease with the font size of 8pt.

The software processes the input image and converts into text format. The image is taken by the user via GPIO pin that are connected to the tactile key by making use of interrupt function [5]. Furthermore, the picture is captured by using raspistill program with sharpness mode to sharpen the image. The resultant image has a .jpg format with a resolution of 3280*2464 pixels.

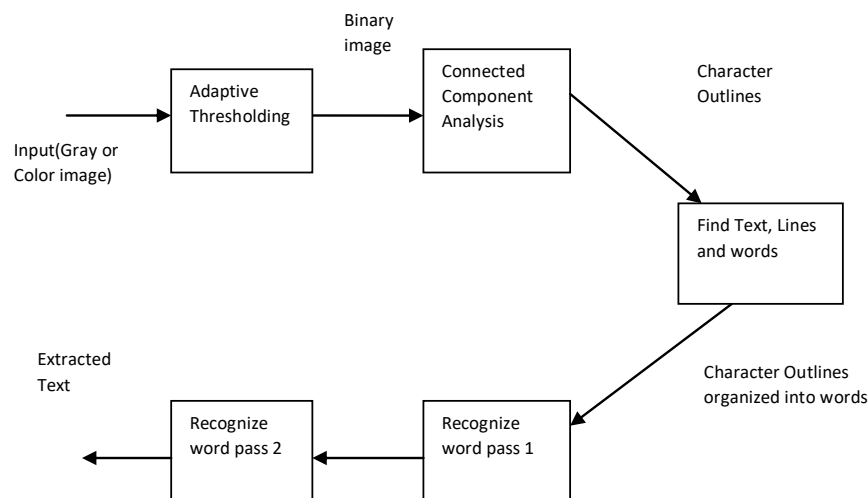


Figure 4 System Design for Image Processing

Figure 4 shows the system design for Image processing module.

Tesseract OCR

B. TTS Correction and Voice Module

In TTS correction and voice module text is converted to speech. The output of OCR is the text, which is stored in a file (speech.txt) [6]. Here, Flite and Espeak software's are used to convert the text to wave format. Finally text.wav can be heard.

Flite and Espeak are open source software's that can be implemented to Raspberry pi, which is available in many

languages. In this research work, English TTS system is used for reading the text.

Flite is a lighter version of Festival built specifically for embedded systems. It has commands that make it easier to use than Festival on the command line[6]. It runs faster than Festival. ESpeak is a compact open source software speech synthesizer for English and other languages, for the operating systems like Linux and Windows. ESpeak uses a "formant synthesis" method. This allows many languages to be provided in a small size. The speech is clear, and can be used at high speeds.

TTS (Text-to-Speech) is a system that can convert input from text into speech. Text-to-Speech in principle consists of two subsystems that are:

a). Text to Phoneme converter

Text to phoneme converter is used to convert the sentence input in a particular language in the form of text into a series of codes that usually represented by the sound of the phoneme codes, its duration and pitch. This section is language dependents.

b). Phoneme to Speech converter

Phoneme to Speech converter will accept input in the form of codes as well as the pitch and duration of phonemes produced by the previous section

4. Design Implementation

a). Import and Initialization

Python's standard library covers a wide range of modules. The voice processing module uses OS package which provides file and process operations, pygame package which provides functions for playing sounds, RPi. GPIO package which provides a class to control the GPIO on a Raspberry Pi, and subprocess package which allows spawning new processes, connect to their input/output/error pipes, and obtain their return codes. The isPause and isStop are variables that will be used for the audio player features. These variables are initialized with a value of False, which means they have not been active. Figure 5 shows the flowchart of the process flow.

b). Setting

We need to import the raspberry pi general purpose input output library from the python library which is in raspberry pi. Then the number is allotted to the GPIO pin in accordance with the breakout board.

c). Main Program

The function of the main program is to provide various commands to retrieve process and convert input image into a sound signal. Various GPIO pins are allotted to control the module operations such as capture, play, pause and stop and to switch off the voice output.

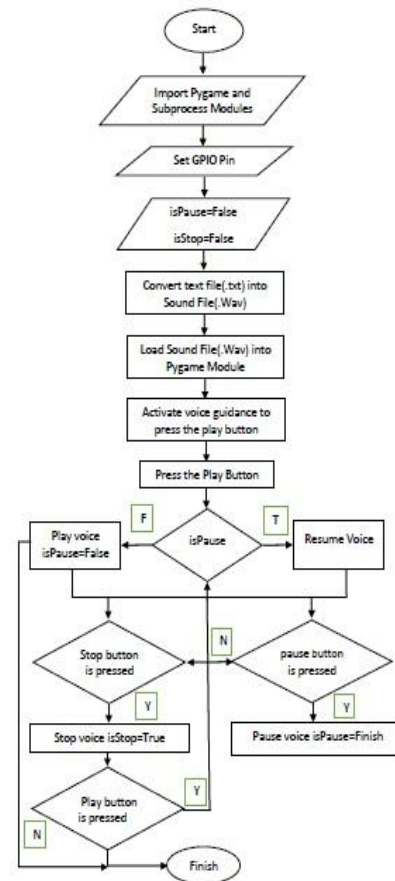


Figure 5 Flow chart of the process flow

We need to import pygame module. Pygame is a module which serves as an audio controller setup which enables us to inculcate various knobs for controlling the playback of the voice.

d). Power Supply Management

Interrupt-based script can be used to enable the best way for soft shutdown feature of the Raspberry Pi power supply switch. Interrupts also improves the efficiency of the code and minimizes load on the CPU compared with while loop. The implementation diagram of power supply switch for Raspberry Pi is shown below. Figure 6 shows the flow chart of the power supply switch.

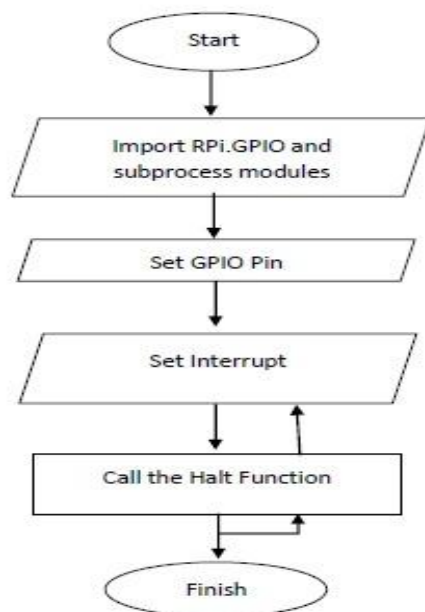


Figure 6 Flow chart of power supply switch

5. RESULT

The testing was done using Raspberry Pi platform with the Following specifications:

- SBU Raspberry Pi 3 900 MHz Quad Code ARM
- Cortex-A7
- Raspberry Pi 8MP Camera Board Module
- Bootable SanDisk Ultra 16GB micro SD Card

From the experimental results it is known that the image processing module has the following restrictions. They are the maximum size of the input image is taken from Magazine. Any input image that uses the block letter fonts will work fine. The minimum font size is 8 point.

We tested each module to see the effectiveness of every Step. Table 1 presents the results of an accuracy testing and average time after enters the image processing module.

Table 1 Results of accuracy testing

Distance in cm	Total words	Errors	Processing Time in seconds	Percentage of Error
30	254	5	54	2.36
15	78	1	23	1.28
18	156	1	38	0.64
25	234	2	47	0.85
26	312	4	54	0.64
20	213	5	52	1.27
24	213	6	45	2.81

From the results of this test, it is showed that the average time of image processing is about one minute or less, depending on the number of the input words that are processed with an average error. This is because the additional state and condition gives better input to the OCR machine.

6. CONCLUSION

In this research work, a Text-to-Speech device for visual impaired people that can change the text image input into sound is implemented. The performance of the device is high enough and it achieves a readability tolerance of less than 2%, with the average time processing less than one minute for various paper and font size. With good lighting, the average error rate from the image processing module is better. This is a portable device and it does not require internet connection, and can be used independently by people with low vision, visual impairment. This device also has a user interface that allows people to interact easily.

REFERENCES

- [1] Global data on Visual impairments, World Health Organization, 2010
- [2] R. Mengko and A. Ayuningtyas. "Indonesian TTS system using syllable concatenation: Speech Optimization" Proc. International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering (ICICI-BME), November 2013, pp. 412-415

[3]ARCHANA A. SHINDE, Dilip Chougule. "Text Pre-processing and Text Segmentation for OCR", International Journal of Computer Science Engineering and Technology", Vol.2 , Issue 1 (2012), pp. 810-812, 2012.

[4]R. Mithe, S. Indalkar and N. Divekar. " Optical Character Recognition" International Journal of Recent Technology and Engineering (IJRTE)", ISSN: 2277-3878, Volume 2, Issue 1(2013).

[5]R. Smith. "An Overview of the Tesseract OCR Engine", USA: Google Inc.

[6]Samuel Thomas, Hema A.Murthy and C. Chandra Sekhar. "Distributed Text to Speech Synthesis for Embedded Systems-An Analysis". Proceedings of the eleventh National Conference on Communications, NCC 2005, pp.1-5.

